NLIDB System TheBest Way to Implement

B.SUJATHA[1] DR.S.VISWANADHA RAJU[2]

Asst. Professor, Dept. of CSE, University College, Hyderabad, Andhra Pradesh, India,
sujatabanothu@gmail.com
Professor & Head in CSE, Dept. of CSE, JNTU College, Karimnagar, Andhra Pradesh, India

Abstract: In Domain Model check it uses data concerning the structure of the domain (supplied by the E-R model).The fields of a node area unit crammed from the data provided by take apart structure and map tables. All the nodes that have over one entity sort within the End-type field area unit appointed a singular entity sort during this part. The domain data that's required is pictured victimization E-R model and generalization hierarchy. Translator The intermediate structure that passes the well-formedness check are a directed tree. The common options area unit our use of history lists for reference, and scrutiny take apart structure of previous sentence for ellipses. Various techniques for illumination area unit victimization Verbs: a number of the verbs of the take apart structure area unit mapped into relations. Accessing the info The NLI system runs on SUN three/60, whereas a sample info for the library domain is made on HCL Horizon-3 machine victimization the relative DBMS, UNIBASE. From the E-R model the entity sorts that area unit constituents of such a relation area unit obtained. An incoming arc into a node indicates that question (or field) cherish the preceding node ought to be generated 1st.

The standard structure of the NLI makes it simply labile to new databases management systems, applications, or alternative Indian languages. We outline node as a node that all the attributes have a worth specific associated a non-value node as a node that an attribute has unspecified value. The question for node is made equally except that it uses the question shaped for node1 conjointly. Keeping the structure of the NLI in such the way that it's simply labile to new applications, databases, or maybe completely different natural languages. Semantic sort Consistency The linguistics sort related to the word is tested for consistency with the linguistics variety of every candidate and people candidates found inconsistent area unit filtered.

Keywords: Question ,Area ,Unit ,Entity ,Node ,Structure ,Word Domain ,Deletion, Sort

I. Introduction

A Modifier has no Head (a) From the previous take apart structure try and notice the top for the modifier (b) Assign new karaka role for the top word (c) Continue with the parsing using the new head for instance, within the following dialogue: ’IEEE’ ke jarnal kOna kOna se hEM? ’IEEE’ -ke journals what -se pres? What area unit the journals of ’IEEE’? ’ACM’ ke kOna kOna se hEM? ’ACM’ -ke what -se pres? which of them of ’ACM’? within the second question there's no head word for the shashti cluster (’ACM’ ke), therefore head word assignment procedure is applied for this shashti cluster with the previous take apart structure because the alternative argument.Recognition The user input is recognized as elliptical at the surface level, if either one or additional of the subsequent cases occur within the parsing phase: (a) No demand word is present (b) a compulsory karaka is absent (c) A Modifier has no head we have a tendency to create associate assumption for surface level deletion that the elliptical fragment corresponds in take apart structure thereto of the parliamentary law and there's no syntax violation within the question. An example illustrating the utilization of domain data in resolution of deep level deletion is given in Section four.2. seven Conclusions Since reference, deletion and alternative phenomena occur often in dialogues involving linguistic communication interfaces to databases, it's necessary to handle them.(If the network model is employed rather than the relative model or another source language is employed, TRANSLATOR module needs to be modified.) The take apart structure is targeted around the verbal teams of the sentence and relies on the karaka (pronounced 'kaarak') relations. Making the intermediate structure totally formed: If the entity sort named by a worth term is unspecified , domain data is employed to search out it. User interaction the utilization of domain data for resolution of deep level deletion has been mentioned History List Formation the most technique for handling intersentential reference uses history list, a record of all objects mentioned within the preceding sentence. The output structure is shown in Third part The module takes the intermediate structure shaped within the previous part as its input, and checks if a sound SQL question will be formed. Algorithms for mapping from take apart structure to intermediate illustration and translating from intermediate illustration to the formal source language

area unit developed. 3 Mapping The clerk module maps the take apart structure to the intermediate illustration whereas the translator module interprets the intermediate illustration to the SQL query. The user input is recognized as elliptical at the deep level, if all the constituents required to create the SQL question can't be obtained from the intermediate illustration of the question. For example, within the following dialogue: Algorithms for mapping from take apart structure to intermediate illustration and translating from intermediate illustration to SQL question area unit developed. • The E-R model has info concerning the relation that every variety of entity has with alternative varieties of entities. The word ideas within the take apart structure area unit mapped into one among the subsequent sorts in intermediate illustration. The verb like one has the data to resolve it as AUTHOR as a result of the corresponding relation WRITE has solely the entity sort AUTHOR in its constituent entity sorts. The entity sort book named by the worth term 'OS concepts' is obtained from this information. Deep level deletion is recognized if all the constituents to create SQL question don't seem to be offered from the intermediate illustration. The take apart structure info is comfortable to handle this sort of deletion. The root node of this tree has the entity sort AUTHOR whereas its kid has the entity sort BOOK. After the entity sort referred by the worth term is found, a replacement node is made for the entity sort. The entity sorts AUTHOR and USER area unit appointed to the sphere End-type of the node. For example within the following dialogue: The entity sort BOOK is appointed to the sphere Ent-type of the node. Case 2: necessary Karaka is Absent (a) within the previous take apart structure establish the word with constant karaka role. Each module builds a illustration of the question in associate acceptable formal structure, till finally the SQL question is made. Node sort illumination a number of the nodes could have over one entity sort within the field End-type. Mapper This module takes the take apart structure and forms the E-R primarily based intermediate illustration. Intermediate illustration and Mapping to Formal question Language: associate E-R model primarily based intermediate illustration is developed. The multidimensional language of the NLI is shown programme, reference and deletion handler, mapper, and translator area unit the process modules whereas the opposite blocks (with rounded corners) are the data sources. 6.5 Domain data in Resolution within the resolution strategies delineate on top of for anaphora and deep level deletion, some quantity of domain data is employed. Translator tables, that provide for every attribute of associate entity sort its corresponding field and record sort within the info schema, area unit used for this purpose. Use previous intermediate illustration, if the previous take apart structure corresponds to current structure three. The detection of anaphora is completed within the programme module and also the resolution is completed later victimization the take apart structure obtained from the programme module. Domain Model check This check uses data concerning the structure of the domain (supplied by the E-R model).The fields of a node area unit crammed from the data provided by take apart structure and map tables. Direct relationship between 2 nodes means within the E-R model there's one relation connecting the entity varieties of the nodes. Domain data Representation: Domain data is captured in associate Entity Relationship (E-R) model for the domain. Entity Attribute Relation price term Quantifier advanced sentences and sentences with multiple verb teams don't seem to be thought-about for mapping in our system. Entity Relationship model is employed to capture the structure of the info. The entity varieties of the nodes connected by a directed arc have a relationship within the E-R model. It relates the entity sorts author and book within the domain. A directed arc connects 2 nodes whose entity sorts have a relationship within the E-R model. Case 1: Demand Word is Absent (a) From the previous take apart structure take the demand word. Thus, the take apart structure contains a mapping between the nominal and also the karaka relations at the side of the ideas of the varied words within the sentence. This check uses the observation that the shashti modifier and its head area unit connected and if the top could be a domain word, the referent of the word ought to have a relation with the head within the domain. The execution of the procedure is as follows: The relation WRITE relates the entity sorts book and author. In the example given in Section four.1, the node cherish the word conception a pair of (node a pair of in . All the nodes that have over one entity sort within the Ent-type field area unit appointed a singular entity sort during this part. Intermediate illustration Intermediate illustration could be a illustration of the user question obtained once analysis. The Paninian programme developed at IIT Kanpur is employed to create the take apart structure of a given question. 3.1 Domain knowledge base data is important for a linguistic communication interface (NLI) to realize an appropriate level of performance. Note that for this instance, the deletion handler is invoked for the second time, the primary present for demand word absence. If a directed arc is from node A to node B, then throughout translation, the question for node A is to be generated 1st to be employed in forming the question of node B later. The Paninian programme is changed to invoke surface level deletion handler in a number of the cases once it finds the sentence to be incomplete. The checks are: • The intermediate structure ought to contain a minimum of one node. The second node shaped, corresponds to the word conception kOna a pair of. The field Ent-type will be crammed by the set of doable entity sorts for a word conception. Like previous approaches we have a tendency to conjointly create use of data 0f info structure (Kumar, 1985) (Kannan, 1987). This part uses domain data, captured within the E-R model, for illumination. This procedure

generates the question by combining the question of every node of the tree.
Deep Level Ellipsis: this is often known and handled throughout intermediate structure construction. All the nodes can have distinctive entity sort once this resolution. The first part forms the nodes of the intermediate structure. A major good thing about victimization linguistic communication is that it shifts onto the system the burden of mediating between 2 views of the information – the manner during which it's hold on (the ”database view”) and also the manner during which associate user worries it (the ”user's view”). The execution of the algorithmic rule for the instance given on top of is as follows: within the 1st FOR loop the nodes of the illustration area unit shaped. The value term will confer with these entity sorts solely. For example, if one shifts from, say, the library domain to air reservation domain, the map tables and E-R model can have to be compelled to be modified. The translator module takes the directed tree and generates the semantically equivalent SQL question. In our system, domain data is pictured in associate E-R model. Absence of a node would indicate that mapping method failing or the NL question was empty. Knowledge illustration in Our System we have a tendency to selected a library data system as associate application space for the example NLI. 6 Reference and deletion To support natural interaction, it's fascinating to permit the utilization of repetition reference and elliptical constructions across sentence sequences. Reference and Ellipsis: the problems of anaphora, ellipsis, and definite noun phrases in natural language interfaces area unit self-addressed. Abstract: For a linguistic communication interface to be helpful, it should settle for anaphora, deletion and alternative means of abbreviating utterances. Value Term Disambiguation: price term illumination may be viewed as breakdown deep level deletion. A set of all unresolved price terms within the take apart structure is additionally obtained. The two nodes with these entity sorts area unit node1 and node2 severally. An unconnected node would, of course, play no role within the overall question. If the word plays a karaka role, we have a tendency to get the linguistics variety of it, from the karaka chart of the most verb. The resolution strategies use constant tests delineate on top of, except that the linguistics sort consistency check will use the linguistics variety of the noun itself. The structure of a node is as shown . Other approaches create use of case frame structures once the syntactical take apart for doing similar things, except these area unit way more data intensive (Rich, 1988). The info machine (Horizon-3: wherever the info resides) is accessed victimization the network to execute the question remotely. 6.1.2 Deep Level deletion Deep level deletion will be detected solely whereas forming the which means of the question. A tool that permits the info administrator (DBA) to enter the E-R model for the domain interactively, will be inbuilt future in order that a site will be simply got wind of. 5.2 Code Generation for Directed Tree The procedure Tree-code-gen is employed to create the SQL question for the tree. If there area unit totally different completely different} sets of relations with different names for a {site} at another site, say, solely the translator tables ought to be modified. The first entry within the table provides the record sort and field for the attribute NAME of node1 (BOOK). The operating of the procedure is delineate victimization the tree shaped for the instance taken. The restricted domain of discourse that's outlined by the info makes it doable to handle these issues in our case.

However, since our take apart structure relies on karaka relations (already known by the parser), the decomposition of the matter is completely different. The child of the basis node (node2) contains a sub-query shaped within the previous iteration. The value term is entered because the default attribute of the node. Only the data sources would like be modified to realize ability to a replacement info, application or alternative Indian languages. The standard structure of the NLI makes it simply labile to new databases management systems, applications, or alternative Indian languages. We outline node as a node that all the attributes have a worth specific associated a non-value node as a node that an attribute has unspecified value. The question for node2 is made equally except that it uses the question shaped for node1 conjointly. Keeping the structure of the NLI in such the way that it's simply labile to new applications, databases, or maybe completely different natural languages. Semantic sort Consistency The linguistics sort related to the word is tested for consistency with the linguistics variety of every candidate and people candidates found inconsistent area unit filtered.

We continue with the instance of the Section four.1. The input to the present procedure is that the directed tree of .The node with no outgoing arcs is selected because the root of the directed tree. Methods for resolution of deep level deletion are: one. So the question of providing access to the present info in simple manner to a novice user is very important.Node1 could be a price node whereas node2 could be a non-value node. For example within the following sentence 'OS concepts' kisane liKI? 'OS concepts' who-ne wrote United Nations agency wrote 'OS concepts'? there's a verb like (to write). The final question shaped is shown . The illustration consists of nodes and directed arcs connecting the nodes.
The final question shaped is shown in . The process of a surface level elliptical fragment involves the subsequent steps: • acknowledge input as elliptical • Get supply in antecedent and construct new full auditory communication one. (b) Continue with the parsing victimisation the word as new karaka. The translator ought to

try and construct these 3 clauses to create the SQL question. Methods to handle these area unit designed and enforced during a example linguistic communication interface (NLI) to databases for Hindi. Ellipsis Handler for every of the on top of 3 cases, the deletion handler is invoked by the programme. But, it's doable to handle them during a restricted domain as in our case. Others Noun teams like "anya kitAbeM" (other books), wherever we've to spot a group of objects, disjoint from the set of objects referred within the previous sentence, are handled.

It is necessary to use the application-specific data, as a result of general linguistic analysis alone can't cause complete formation of the which means of the question. User Interaction If the on top of listed filters fail to offer the referent, we have a tendency to raise the user to pick it from the set of candidates employing a menu. Checking for grammaticalness: A check is created to ascertain whether or not the intermediate illustration obtained is well-formed. History list is made once the anaphora resolution is over for the present sentence. Surface Level Ellipsis: this is often detected and handled at syntactical level by the Paninian programme.That programme has been extended to handle deletion. Taxonomies with a hierarchy of sorts and subtypes will be accustomed prohibit the quantity of entities and relations. While forming the question of node2 (AUTHOR) the question hooked up to node1 is employed. They represent the bottom domain data during a compact manner. This info is employed to construct the question for it. Blocks on top of layer a pair of and below layer one will be modified if a special application domain is chosen. The ensuing SQL question might not be optimum from the info retrieval view-point. The programme is changed in order that rather than failing for the on top of cases, it invokes associate deletion handler. In the second FOR loop, the relation WRITE, cherish the verb like one is entered during a set of relations encountered. In this variety of reference solely the objects already gift within the colloquial context area unit known. Besides the karaka relations, the sentence could contain the non-karaka relations like those contributed by the adjective relations, purpose and relative words. A linguistic communication Interface (NLI) operational in conjunction with a menu-based system could be a sensible candidate for this purpose. So, a directed arc from node1 to node2 is made and it's tagged by the relation WRITE. The programme continues with its task once deletion handler returns management. It is assumed that quantifier scoping isn't necessary for the input question which all the relationships between nodes area unit direct. Thus wherever scopes of quantifiers area unit necessary, the NLI fails to provide associate acceptable formal question. The karaka relations area unit used for the illumination of word senses. 6.1 deletion an issue is termed elliptic if one or additional of its constituents area unit omitted. (b) Continue parsing victimization the new demand word. This ambiguity will be resolved by victimization the techniques delineate within the previous section. The E-R model consists of entities, attributes and relations between entities. The answer file is processed to indicate answers in appropriate chunks (e.g., implementing numeric quantifiers mentioned in user's query) and to tell the user just in case of NULL answers. Reference and deletion area unit exhausting issues during a general context. An attribute is joined to its entity set by associate purposeless edge. Finally, directed arcs connecting the nodes area unit shaped. Use domain data a pair of. The parallelogram represents associate entity set and also the oval represents associate attribute. The direction of the arcs is from a toddler node to its parent node. This clause tells that record sort the fields area unit to return from. A question optimizer will be used on the generated query before passing it to the DBMS. The E-R diagram for the library domain is shown in . Directed Arc Construction The nodes area unit connected by directed arcs during this part. We area unit able to create use of linguistics options in karaka charts themselves that area unit a part of the programme. The 'from' clause is additionally needed for a sound question. It is necessary for associate NLI to handle deletion for brevity in communication. Basically, associate SQL question consists of clauses, every of that is preceded by a keyword. They are joined to their constituent entity sets by purposeless edges. All the relationships shown within the area unit of sort many-many. The 'select' clause introduces each question. It is pictured by an acceptable organization. Here, solely an instantaneous relationship between 2 varieties of entities is assumed. These tables conjointly contain the be part of conditions necessary to attach 2 record sorts. 6.3.1 Anaphora The term 'anaphora' refers to reflexive pronouns, general pronouns, definite noun phrases, etc. A filter takes a group of doable candidates as its input and provides those candidates that pass the check as its output. For example, there's associate ISA link from manuals to books as a result of a manual could be a book. Anaphora resolution involves finding referents of those during a discourse which can include over one sentence. While taking a specific karaka role as default referent could be a sensible choice, it doesn't work continually. • If ambiguity is gift even once the applying of the on top of techniques then user interaction is required. • each node ought to have a minimum of one incoming arc and at the most one outgoing arc. Introduction The community of potential data system users is growing quickly with advances in hardware and computer code technology The procedure Node-gen is termed 1st with node1 (BOOK) as its argument. The approach adopted by US has many common options also as variations with alternative approaches (Grosz et al., 1990) (Carbonell, 1983) (Bolc,

Jarke, 1986) (Jones,1988). We continue with the instance of Section four.1 to indicate the functioning of the algorithmic rule. 6.3.2 Definite Noun Phrases The detection of definite noun phrases is completed by victimization the determiner employed in the phrase. The record sorts rt2 and rt1 area unit joined on the fields accnum and accno severally. The map tables offer the data necessary for this mapping. The relations area unit hold on to be used within the second part. Some of the numerous area unitas wherever mechanization is being done quickly are railway reservations, management info systems, library services etc. A sample endure an issue is shown in Appendix A. The tests developed for handling inter-sentential anaphora will be used for intra-sentential case also. 6.3 Reference we have a tendency to think about the case of intersentential repetition reference solely. Second part The second part consists of 3 stages for the subsequent 3 tasks. It is supported Entity-Relationship model. The modifiers of the word area unit 'databases'. This is used as associate info supply cut loose the process modules.

CONCLUSION: The standard structure of the NLI makes it merely labile to new databases management systems, applications, or different Indian languages. We define node as a node that everyone the attributes have a value specific associated a non-value node as a node that associate degree attribute has one worth. The question for node is formed equally except that it uses the question formed for node1 together. Keeping the structure of the NLI in such the approach that it's merely labile to new applications, databases, or even fully totally different natural languages. Semantic type Consistency The linguistics type associated with the word is tested for consistency with the linguistics form of each candidate and folks candidates found inconsistent unit of measurement filtered.

REFERENCES

[1] G. R. Rocher Silva, *Traducción de Queries en Prolog a SQL*, BE thesis, Escuela de Ingeniería, Universidad de las Américas- Puebla, 1999.

[2] J. Rojas, J. Torres, A Survey in Natural Language Databases Interfaces, *8vo. Congreso Intern. de Investigación en Ciencias Computa- cionales*, Inst. Tecnol. de Colima, Mexico 2001.

[3] M. Ruiz, A. Diekema, P. Sheridan, *CINDOR Conceptual Interlingua Document Retrieval:TREC-8 Evaluation,* MNISTextWise Labs, Syracuse, NY 13202.

[4] R.J.H. Scha, Philips Question Answering System PHILIQA1, *SIGART Newsletter*, No. 61. ACM, New York, 1977.

[5] V. Sethi, *Natural Language Interfaces to Databases: MIS Impact, and a Survey of Their Use and Importance,*University of Pittsburgh, PA, USA.

[6] DCC, *SiMBaDD Sistema Manejador de Basesde Datos Distribuidas*, CENIDET, Cuernavaca, Mexico, www.sdcenidet. com.mx/simbadd.

[7] D. Stumberger, B. Ballard, Semantic Acquisition in TELI, *24th Annual Meeting of ACL,* New York, 1986.

[8] M. Templeton and J. Burger, Problems in Natural Language Interface to DBMS with Examples from EUFID, *1st Conference on Applied Natural Language Processing*, Santa Monica, California, 1983.

[9] B.H. Thompson and F.B. Thompson, Introducing ASK, A Simple Knowledgeable System, *1st Conference on Applied NaturalLanguage Processing*, Santa Monica, CA, 1983.

[10] B.H. Thompson, C. Raymond, and J. Money, *Automatic Construction of Semantic Lexicons for Learning Natural Language Interfaces,* Stanford Univ., University of Texas.

[11] Trinzic Corporation, Bethesda, MD *INTELLECT–Natural Language System.* (commercial triptych).

[12] D.L. Waltz, An English Language Question Answering System for a Large Relational Database, *Comm. of the ACM,* 21(7), 1978.

[13] D. Warren and F. Pereira, An Efficient Easily Adaptable System for Interpreting Natural Lan- guage Queries, *Comp.Linguistics,* 8 (3-4), 1978.

[14] A. W. Penn, *DB Valet: A natural language database interface*, MS thesis, Univ. of Lousville, 2000, www.louisville.edu.

[15] W.A. Woods, Procedural Semantics for a Question-Answering Machine *Fall Joint Computer Conference*, NY, 1968. AFIPS.

[16] J. Allen. Recognizing Intentions from Natural Language Utterances. In M. Brady and R.C. Berwick,editors, Computational Models of Discourse, chapter 2,pages 107–166. MIT Press, Cambridge, Massachusetts, 1983.

[17] H. Alshawi. The Core Language Engine. MIT Press, Cambridge, Massachusetts, 1992.

[18] H. Alshawi, D. Carter, R. Crouch, S. Pulman, M.Rayner, and A. Smith. CLARE – A Contextual Reasoning and Cooperative Response Framework for the Core Language Engine. Final report, SRI International, December 1992.

[19] I. Androutsopoulos. Interfacing a Natural Language Front-End to a Relational Database (MSc thesis). Technical paper 11, Department of Artificial Intelligence,University of Edinburgh, 1993.

[20] I. Androutsopoulos, G. Ritchie, and P. Thanisch. An Efficient and Portable Natural Language Query Interface for Relational Databases. In P.W. Chung, G. Lovegrove,and M. Ali, editors, Proceedings of the 6th International

Conference on Industrial & Engineering Applications of Artificial Intelligence and Expert Systems, Edinburgh, U.K., pages 327–330. Gordon and Breach Publishers Inc., Langhorne, PA, U.S.A., June 1993. ISBN 2–88124–604–4.

[21] P. Auxerre. MASQUE Modular Answering System for Queries in English - Programmer's Manual. Technical Report AIAI/SR/11, Artificial Intelligence Applications Institute, University of Edinburgh, March 1986.

[22] P. Auxerre and R. Inder. MASQUE Modular Answering System for Queries in English - User's Manual. Technical Report AIAI/SR/10, Artificial Intelligence Applications Institute, University of Edinburgh, June 1986.

[23] B. Ballard and D. Stumberger. Semantic Acquisition in TELI. In Proceedings of the 24th Annual Meeting of ACL, New York, pages 20–29, 1986.

[24] B.W. Ballard. The Syntax and Semantics of User-Defined Modifiers in a Transportable Natural Language Processor. In Proceedings of the 22nd Annual Meeting of ACL, Stanford, California, pages 52–56, 1984.

[25] B.W. Ballard, J.C. Lusth, and N.L. Tinkham. LDC-1: A Transportable, Knowledgebased Natural Language Processor for Office Environments. ACM Transactions on Office Information Systems, 2(1):1–25, January 1984.

[26] M. Bates, M.G. Moser, and D. Stallard. The IRUS transportable natural language database interface. In L. Kerschberg, editor, Expert Database Systems, pages 617–630. Benjamin/Cummings, Menlo Park, CA., 1986.